



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### The tyrosinase-related protein-1 gene has a structure and promoter sequence very different from tyrosinase

**Citation for published version:**

Jackson, I, Chambers, DM, Budd, PS & Johnson, R 1991, 'The tyrosinase-related protein-1 gene has a structure and promoter sequence very different from tyrosinase', *Nucleic Acids Research*, vol. 19, no. 14, pp. 3799-804. <<http://nar.oxfordjournals.org/content/19/14/3799.long>>

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Nucleic Acids Research

**Publisher Rights Statement:**

Copyright 1991 Oxford University Press

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# The tyrosinase-related protein-1 gene has a structure and promoter sequence very different from tyrosinase

I.J.Jackson, D.M.Chambers, P.S.Budd and R.Johnson

MRC Human Genetics Unit, Western General Hospital, Crewe Rd, Edinburgh EH4 2XU, UK

Received June 6, 1991; Revised and Accepted June 21, 1991

EMBL accession no. X59513

## ABSTRACT

**We have determined the exon structure of the mouse tyrosinase-related protein-1 (TRP-1) gene. The gene is only 15kb in length, but contains seven introns, in contrast to the tyrosinase gene which is almost 100kb long with only four introns. Only two introns are located in homologous positions in both genes. Intron I of TRP-1 has three alternative 5' splice sites clustered within 21bp, which all splice to the same 3' site. Intron V has a very unusual 5' splice site, which has the dinucleotide GC rather than the conventional GT. We show that as little as 370bp of 5'-flanking DNA is sufficient to direct cell-specific expression of the chloramphenicol acetyl transferase gene. The flanking DNA of TRP-1, unlike tyrosinase, does not contain a TATA box or a CCAAT box. Both mouse genes, however, share an 11bp sequence, also found in human tyrosinase, which we suggest may be a melanocyte-specific promoter element.**

## INTRODUCTION

Synthesis of melanin by melanocytes requires a number of enzymes (1). Key among these is tyrosinase, which catalyses the conversion of tyrosine to dihydroxyphenylalanine (DOPA) and DOPA to dopaquinone (2). A number of other enzyme functions have been identified, such as dopachrome tautomerase and dihydroxyindole reductase, which are probably not essential for pigmentation but which accelerate reactions thought to occur spontaneously (2–5).

Tyrosinase cDNA has been cloned, both from mouse and human (6–9). It maps in the mouse to the *albino* locus on chromosome 7 (10), and specific mutations have been identified in mouse and human albino individuals (11–19). A second cDNA has been isolated which encodes a protein clearly related to tyrosinase, with about 52% amino acid identity (20). This gene, called tyrosinase related protein -1 (TRP-1), maps to the murine *brown* locus on chromosome 4 (21). Null mutations in this gene result in the production of brown rather than black eumelanin (22, 23). A third member of this gene family, TRP-2, has been identified and it encodes a protein equally related to the other two (21 and IJJ, unpublished).

The mouse and human tyrosinase genes have been cloned, their intron structure determined, the end of their transcripts mapped, and the regulatory regions 5' of the transcription units sequenced and analysed (24–29). We have performed a similar analysis

on the mouse TRP-1 gene. Based on the clear evolutionary relationship between the two genes we expected to find strong similarities between the gene structures. However, we have found that TRP-1 has seven introns, compared to four in tyrosinase, only two of which are in conserved positions in the two genes. One of these introns in TRP-1 has a highly unusual 5' splice site, which nevertheless is used accurately. Although the 5' end of the tyrosinase gene contains a number of recognised cis-acting regulatory elements, including TATA boxes, CCAAT boxes and several palindromic and direct repeat sequences, no such elements are present in a region of several hundred bases pairs which flanks the TRP-1 transcriptional start site, despite the fact that this region of DNA is sufficient to specifically direct synthesis of chloramphenicol acetyl transferase (CAT) in melanoma cells. We identify an 11bp sequence which is found between 40 and 140 upstream of the multiple transcriptional start sites of TRP-1 and about 90 and 140bp upstream of the major transcriptional start sites of mouse and human tyrosinase respectively, which we suggest may be a melanocyte specific cis-acting element.

## MATERIALS AND METHODS

### Library screening, clone isolation, subcloning and sequencing

The bacteriophage EMBL3, partial Sau3A library of mouse strain 129 genomic DNA was a gift from Lisa Stubbs (ICRF, London). The library was screened by plating  $2.6 \times 10^5$  pfu on *E.coli* strain Q358 at a density of about 160 pfu/cm<sup>2</sup>. Replicas of the plate were made by overlaying with nylon membrane (Hybond-N; Amersham) and fixed by UV irradiation and baking. The filters were hybridised in 0.5M sodium phosphate, pH 7.2, 7% SDS (30) with fragments of the TRP-1 cDNA, pMT4 (20), labelled with <sup>32</sup>P using DNA polymerase I in a random-primed labelling kit (31)(Boehringer-Mannheim). Positive phage were picked and rescreened to single plaques. DNA was purified from the phage, restriction mapped with a number of enzymes, and suitable fragments subcloned into pBluescribe (Stratagene). Exon and intron/exon junction sequences were obtained using Sequenase kits (USB) on double-stranded DNA with primers flanking the polylinker (as supplied), or primers as described below.

### Oligonucleotide primers

All oligonucleotides were synthesised on ABI 381A or 391 DNA synthesisers, deprotected overnight at 65°C in 30% ammonium hydroxide and precipitated from 0.3M sodium acetate with 70%

ethanol. Primers for sequencing the exons were designed from the published cDNA sequence, as previously described (20, 22). Further primers were made according to sequences determined 5' of the gene and within intron I to allow confirmation and extension of sequences.

### Polymerase chain reaction

Alternative splice sites were analysed in cDNA from total RNA prepared from neonatal mouse skin as described. The polymerase chain reaction was performed using primers CAATTAACAGCTGGCATC (bases -174 to -157) and TCTCGTGGAACTGAGCC (complement of bases 85 to 68) as described previously (22, 32), denaturing initially at 94°C for 2 minutes, followed by 30 cycles of 94°C for 15 seconds, 55°C for 15 seconds and 72°C for 30 seconds. The products were analysed on 4% Nuseive agarose gels, isolated from the gel using GeneClean (BIO 101, La Jolla, CA), and sequenced directly with the amplifying primers in 10% DMSO according to the method of Winship (33).

### Primer extension

The 5' ends of the transcripts were determined by extension of an oligonucleotide primer GGAAGGTTTCTCTGCTGA, covering bases -97 to -114 of exon 1, using RNA prepared from B16 melanoma cells. The primer was 5'-end labelled with <sup>32</sup>P and purified after electrophoresis in polyacrylamide/urea gel. 16,000 cpm were annealed to 20 µg total RNA at 70°C and cooled to room temperature. The primer was extended with AMV reverse transcriptase in 10mM Tris, 10mM MgCl<sub>2</sub>, 1mM DTT, 1mM deoxynucleotide triphosphates for 1 hour at 42°C. The products were analysed by electrophoresis on 6% polyacrylamide/urea gels and visualised by autoradiography.

### Melanoma cell culture and CAT assay

Fragments from the 5' end of the TRP-1 gene were cloned into the chloramphenicol acetyl transferase (CAT) expression vector pBLCAT3 (34). Both fragments terminated at their 3' end at the PvuII site at base -164, and at the 5' end at either the HindIII site at -642, or the XbaI site at approximately -1600. B16 melanoma cells were normally maintained in MOPS-L medium, which consists of HAMS F10 medium supplemented with glutamine buffered by 187 mM MOPS and 84 mM NaHCO<sub>3</sub>. This gives a slightly acidic pH which minimises pigment production. Following transfection the cells were transferred to DMEM, which has a slightly alkaline pH and stimulates melanin synthesis. Cells were grown in a 5% CO<sub>2</sub> atmosphere. DNA constructs were introduced into cells using lipofectin (BRL, Gaithersburg, MD) according to the manufacturer's instructions, the DNA was allowed to express for 48 hrs and the cells harvested. The CAT reaction was assayed as described using <sup>14</sup>C-acetyl CoA (35); the products being separated by thin-layer chromatography, autoradiographed, and quantitated by scraping the spots and liquid scintillation counting.

## RESULTS

### Isolation and gene structure of the TRP-1 gene

A library of mouse DNA, from strain 129, cloned into EMBL4 was screened with probes from the TRP-1 cDNA, pMT4. Six strongly hybridising phage were identified and isolated. Restriction mapping and hybridisation to specific oligonucleotides showed that these fell into two classes. Two clones did not hybridise well to any oligonucleotide, and the region of

hybridisation to the cDNA was restricted to about 1.5 kb. We believe that these clones represent a second locus, previously described (21), which maps apart from TRP-1, and which is probably a processed pseudogene. The remaining 4 clones form an overlapping set, spanning about 38 kb of mouse DNA.

Restriction mapping, subcloning of fragments, sequencing of exons using specific oligonucleotides and PCR between oligonucleotides on neighbouring exons has allowed us to derive the map of the TRP-1 gene in Figure 1. The gene is divided into eight exons. The first exon is entirely 5' non-coding sequence, the second is 5' non-coding and protein coding and the eighth is both coding and 3' non-coding. All the introns are fairly small, ranging from 0.55 to 2.75kb, and the exons are fairly uniform in length; with the exception of the final, 1kb exon, they range between 147 and 470 bp. The overall length of the TRP-1 transcription unit is thus less than 15kb. Table I summarizes these sizes.

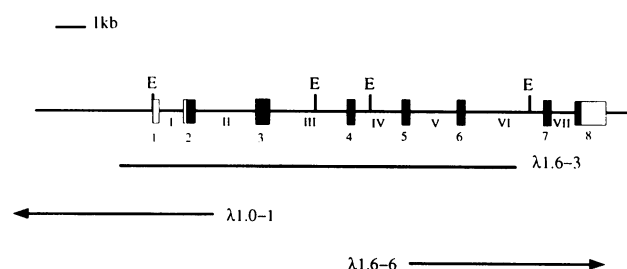
We have sequenced the whole of intron I, and 5' flanking DNA extending for several hundred base pairs 5' of the gene. Figure 2 shows this sequence, on which is marked the transcriptional start sites, and the alternative splice sites of intron I (see below). We have identified, by search of the sequence databases, a highly diverged B2 repeat element in this intron, between approximately bases 755 and 925 (Figure 2). The B2 sequence has equal similarity (65–70%) with both the rat and mouse repeats and may represent an ancestral member of this family.

### Splicing of the TRP-1 transcript

When we sequenced a new TRP-1 cDNA clone we found an apparent insertion of 6 base pairs at the exons 1 and 2 boundary, when compared to the sequence from Shibahara *et al* (20).

**Table 1.** Size and locations of introns and exons. Numbering is according to Shibahara *et al*. The 4 transcriptional starts and 3 alternative splices results in potentially 12 different first exons, these are shown in Figure 1.

exon	size (bp)	position	intron	size (kb)
1	87 to 208	see Fig 1	I	0.57 to 0.59
2	470	-85 to +385	II	2.0
3	323	386–708	III	2.75
4	205	709–913	IV	1.6
5	168	914–1081	V	1.7
6	180	1082–1261	VI	2.45
7	147	1262–1408	VII	0.8
8	1041	1409–2449		



**Figure 1.** The structure of the TRP-1 gene. Exons are numbered 1 to 8 (Arabic), and introns I to VII (Roman). Open boxes represent 5' and 3' untranslated regions of the mRNA, filled boxes are the protein-encoding regions and the line is flanking and intronic DNA. Below the gene are shown three  $\lambda$  recombinant clones, from which the gene structure was determined. E = EcoRI sites.

Inspection of the genomic sequence in this region shows that these six bases were derived from the 5' end of intron I, by use of a second consensus splice site. To further examine splicing of intron I we used the polymerase chain reaction to amplify cDNA fragments spanning the junction of exons 1 and 2. High resolution agarose gels reveal at least 3 amplification products which were isolated and sequenced directly. Three alternative 5' splice sites were found; the one resulting in the cDNA described by Shibahara *et al.*, one 6bp 3' of this, as found in our cDNA and a third a further 15bp downstream. The sites are shown in Figure 2 as IA, IB and IC. All three conform to the 5' splice consensus sequence (see Table 2) and all are joined to the same 3' site. The three PCR products were of approximately equal abundance suggesting that all three sites are used approximately equally.

Table 2 lists the sequences at the nine 5' and seven 3' sites of all seven introns. The exact splice site, where any ambiguity exists, has been placed so that the sequences conform to the established consensus, which is that introns invariably begin with the dinucleotide GT and end with AG (but see below); coupled with a comparison to the cDNA sequence, this has allowed exact

positioning of all the introns. Intron V is an example of a very rare exception to the rule, in which the 5' GT is replaced by GC. Other than these strongly conserved dinucleotide pairs, splice sites generally have a loose consensus (36–38). At the 5' end, there is an A or C at –3 in 69% of cases examined, at –2 is A in 60% of sites, and at –1 is G (79%). The 3' consensus consists of runs of pyrimidines in the 10 to 12 bases before the conserved AG. All the TRP-1 splice sites conform to these patterns.

### The TRP-1 Promoter

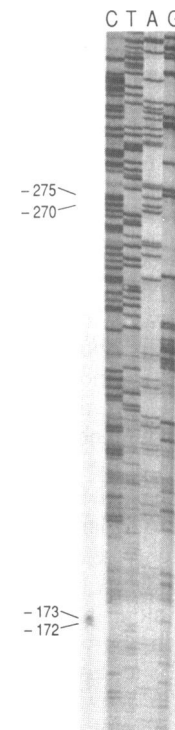
We have determined the 5' end of the TRP-1 transcripts by extension of an oligonucleotide primer using reverse transcriptase on RNA from melanoma cells. Figure 3 shows the result of such an assay. At least 4 specific fragments are consistently observed, extending the primer by 56, 57, 153 and 157 nucleotides. This is consistent with 5' ends at –172, –173, –269 and –273,

**Table 2.** Sequences surrounding the 5' and 3' splice sites of the 7 TRP-1 introns, including the 3 alternative 5' splices of intron I.

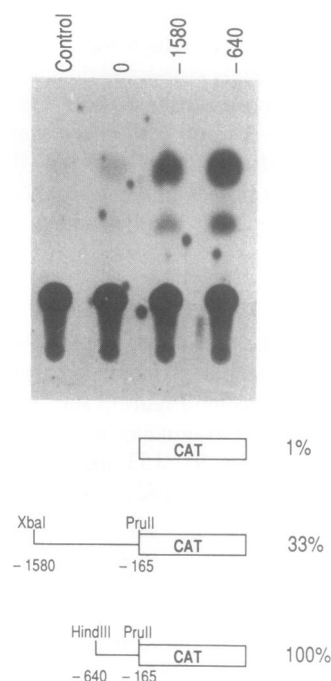
Intron	5' splice	3' splice
IA	ATTCATG/GTACTGGT.....	
IB	GGTACTG/GTGAGCAG.....	
IC	CTCTGTG/GTGGGTAC.....	.....CTGTTTTTCAG/CTGTA
II	CTCACAG/GTGAGTAG.....	.....CATGTTTAAAG/TCAGG
III	CATGCAG/GTATGCTG.....	.....CTATGATCTAG/GAGAT
IV	TGTAACA/GTAAGACC.....	.....TTCCTCCCAG/GCACT
V	GTGGAAG/GCAAGTAA.....	.....TTCAAATGTAG/GTTAC
VI	AACGCCG/GTGAGATC.....	.....GAATATTTTAG/ATATT
VII	TGGCCAG/GTCGGTTT.....	.....CTTCAAATAG/GTCAG

1 AAGCTTTGTAGAGTAATCATGTATTCCAACTCAGGCTTACATTTGAATGTTGGCTACAT 60  
 61 ATGTATGAGTTTTCAACTTCCAGGAGAAAACGTCTCTTTAAAAGAGAAACACAAAAGCT 120  
 121 AACAGAAAATACAAAGTGTGACATTGGCCTTAGTTCGACCAAGAAAGCAATTCATCTTGT 180  
 181 TCTTCCTTTGTGTATACAGATAAGAAAAATAAATCACTACAACGAAAGCAAAATCTCT 240  
 241 TCAGCGCTCTTAATACATCTTCAAAATCAGTGTGTCTGACCTTTTCTTAAGACTTTAAC 300  
 301 ATCACAGGAAACAGTGGGGAGGAGTCAATGCTAGTAGTTAAAGGCAGGAGA 360  
 361 ATTCACCTGGTGTGAGAGGATTAGTGAGAGCTGGAAGAGAGGACCGCCCTCCAGTG 420  
 421 TGAGGAATCTGGCTTGGGATTACTGTCTGGCAGAAAATCTCTCGGGCAATTAACAGCT 480  
 481 GGCATCAGGGGAAAGCAGACATCCAACTAGCTCTGAAGGAGATCAGCAGAGAAA 540  
 541 CCTTCCAGGGATTCTATGTTACTGTGTGAGCAGCTCTGTGGTACCCCTTGTGACAAA 600  
 601 CTCTAGGAACATGAAGGAGATTGCTTGTCTATAAACCCTGTTCTCTATTCTCTTCATT 660  
 661 CCATGGTTAACTATTACTATGGTAGTCACCACTAGTGGATGCTTTTGGTAAATGACATC 720  
 721 TATGGAAGTCTTTTGGATCAGGGTGATCTTTTATGTATGTGTATGTGATGATATG 780  
 781 GGTGCACGAGAGCAGGTGCCAGATTCTCAAGGAGGCTTCAGTTACAAGGAGTGGGAG 840  
 841 TGATCTGATGTGGTTGCAAGGCACTGAAGTCAGTCTCTGTGAAGGCACTCTATGCTCC 900  
 901 TTACCACTGTGCTTCTCCCGAGCCCAAGATAGTATTCTTATGGGTAGAAATTTAAATA 960  
 961 AGAACTCAAGACAGGAGAGTGAGTTCTGTCTATGCTATTATGCCTGCAGATATT 1020  
 1021 AAAGGTGAATAATTATTTGACTATTGTTTAGAAATGTTGTTTCACATGAAAGATTCCAT 1080  
 1081 TTCCGAGTGGGTTGAAAGTATGCAAAAGAACTTTTGCAACTCTGTTTTCCTTTCTG 1140  
 1141 TTTTTCAGCTGTATTTTCACTGAGCACCCTGTCTTCTCCATGCAAGAGCAGCATAGG 1200  
 1201 AGACCTGTGTTCTGAACCTCTGCTTCGAGAGAATG 1236

**Figure 2.** The sequence of the 5' flanking DNA, the first exon, first intron and exon 2 up to the translational initiation codon. Exon sequences are underlined. Splice sites are shown by vertical arrows and transcriptional initiation sites by bent arrows, but note our data do not allow precise nucleotide assignment of the initiation sites. The 11mer motif shared by TRP-1 and tyrosinase is boxed. The sequence is deposited in the EMBL sequence database with accession number X59513.



**Figure 3.** Oligonucleotide primer extension to determine the 5' ends of the TRP-1 transcripts. The four extended products are labelled according to the positions of their initiation. The adjacent sequence ladder is determined using the same oligonucleotide to prime a sequence reaction.

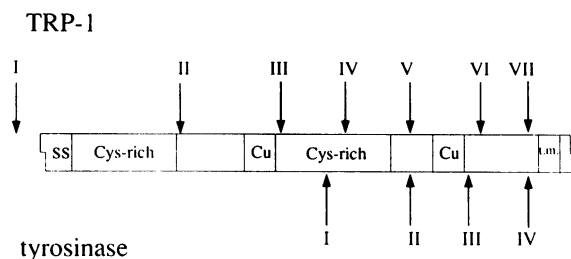


**Figure 4.** Assays of CAT expression of constructs in B16 melanoma cells. The autoradiograph of thin-layer chromatography shows control (non-transfected) cells, and cells transfected with pBLCAT3 alone (0), pBLCAT3 plus TRP-1 DNA to the XbaI site (–1580) and with TRP-1 DNA to the HindIII site (–640). The line figure shows a schematised view of the constructs, and the relative expression of CAT from each.

although it is not possible to assign the ends to specific nucleotides given the fairly broad bands seen on the gel. (When referring to positions 5' in the sequence we use the numbering of Shibahara *et al* (20), in which the base before the translational initiation codon is –1, and the smallest spliced form is used). Sizing of the fragments is by alignment with a sequencing ladder obtained using the same primer on a subclone of the region and run in adjacent tracks in Figure 3.

In Figure 2 we show the sequence of the DNA flanking the TRP-1 gene, indicating these 4 start sites, and extending to the HindIII site at –643. In this region there are no good matches to any described consensus cis-acting sequences, in particular there are no TATA or CCAAT boxes. The absence of a TATA box and thus lack of a TFIID binding site is consistent with the heterogeneous 5' ends of TRP-1 mRNA.

To ensure that these 5' sequences nevertheless were able to direct specific transcription of the TRP-1 gene, we placed them upstream of the CAT gene in the vector pBLCAT3 (34), and introduced the constructs into melanoma and other cells. Figure 4 is the result of CAT activity assays on B16 melanoma cells transiently transfected with DNA from various CAT constructs. The promoterless vector alone has measurable but very low activity in the assay. When approximately 1300bp of DNA 5' of TRP-1, up to the XbaI site, are ligated upstream of a CAT gene, a 30-fold increase in activity is seen after transfection. When a smaller piece of DNA is used, to the HindIII site (the sequence in Figure 2), a greater transcriptional activity is consistently seen. Removing the ~700 bp between the XbaI and HindIII sites results in a further 3-fold increase in CAT activity, suggesting the presence of a negative transcriptional element in this interval. The same short construct, introduced into mouse



**Figure 5.** Schematic representation of the protein domains of tyrosinase and TRP-1. SS = signal sequence; Cys-rich = cysteine-rich domains; Cu = putative copper-binding sites; t.m. = putative transmembrane region. The arrows indicate the positions of introns I to VII of TRP-1 and I to IV of tyrosinase. Introns V and II and VII and IV are in conserved positions in the two genes, as shown.

L-cells or NIH 3T3 cells gave only between 1.5 and 5.5% CAT expression as compared to its activity in B16 cells (data not shown); indicating the melanoma specificity of this promoter element.

## DISCUSSION

### An unusual 5' splice site

Whilst most of the bases which make up the consensus sequences determining the position of an RNA splice are not individually essential, the dinucleotide GT at the 5' end and AG at the 3' end of the intron are almost invariant. Shapiro and coworkers have extracted splice sites from the sequence databases and compiled updated consensus sequences and they describe a number of splice sequences which do not match the GT–AG rule (37,38). Many of these can be eliminated as errors of data entry or interpretation (see 55 for a discussion). Virtually all of the remaining non-GT 5' sites, like TRP-1 intron V, have GT replaced by GC. Aebi *et al* (39) demonstrated, by assaying for splice site function *in vitro*, that the only alternative to GT which would direct an accurate splice was GC, although the reaction proceeded more slowly.

If this intron in TRP-1 pre-mRNA is removed at a slower rate, it nevertheless does not seem to affect the fidelity of splicing. Primers spanning potential alternative splices in the PCR on neonatal skin cDNA should permit any such variants to be detected, but none are (data not shown). Any alternative splices affected by the unusual intron V 5' end must be very rare, if they exist at all, to escape detection in this sensitive assay.

### TRP-1 gene structure is very different from tyrosinase

The strong sequence similarity between TRP-1 and tyrosinase indicates a common evolutionary ancestor. We have compared the positions of introns in the TRP-1 gene, with their positions in the tyrosinase gene. The sequence similarity between the two proteins is such that intron locations can be assigned unequivocally to particular homologous codons. The TRP-1 gene has an intron in the 5' untranslated region, which tyrosinase does not. Within the protein coding sequence, TRP-1 has 6 introns and tyrosinase has 4, but only 2 of these are present in homologous positions: introns V and VII of TRP-1 and introns II and IV of tyrosinase. Both pairs are in codons which can be unequivocally identified as homologous, and both are in the same positions within the codon. The unusual GC 5' splice site is in TRP-1 intron V, although the homologous tyrosinase intron II has a conventional GT site.

The tyrosinase-related family of proteins have five recognisable protein domains (1); two copper-binding sites with similarity to sites in lower eukaryote and prokaryote tyrosinases (1, 7, 40, 41), two cysteine-rich domains of unknown function, and a transmembrane domain by which the enzymes are anchored to the inner face of the melanosomal membrane. Figure 5 illustrates these domains, and indicates the locations of the introns in tyrosinase and TRP-1. Most models of gene evolution have been based on the idea that exons are the basic evolutionary unit, and exon shuffling produces a diversity of multi-exon genes. An exon, in this model, is or was at one time capable of encoding a discrete protein domain. There are, however, only very few clear examples where exons are reused in different contexts.

Whilst the introns of tyrosinase and TRP-1 do separate protein domains, we need to explain why most of the introns are in different locations in the two genes. The usual assumption is that a new gene becomes assembled from a set of exons, which may later be duplicated and introns are gradually lost at random in the two related genes. This may be true of the evolution of the tyrosinase/TRP-1 family, although it predicts that some of the ancestral exons were be rather small, including one encoding a domain of only 18 and one of only 11 amino acids. Whilst such small exons could have existed it is worth considering alternatives.

One alternative is that introns have been gained in evolution as well as lost. The view put forward compellingly by Cavalier-Smith (42) is that introns were introduced into most eukaryotic genomes very early in their evolution, about 1000 million years ago, as self-splicing retrotransposons which degenerated following the evolution of spliceosomal processing. The suggestion is that insertion occurred over a period of several hundreds of millions of years. Much of the evidence accumulated over recent years which makes such a model compelling has been reviewed by Rogers. (43–46). When genes encoding the same protein (orthologous genes) are compared between vertebrate species introns are almost always found in the same positions. However, when orthologous genes from invertebrates, such as *Drosophila*, and vertebrates are compared they frequently have different intron locations (47). Comparison can also be made within genes where there has been an internal duplication to result in two related halves. In many such cases, for example the *mdr-1* gene (48), the intron positions in the two halves are different.

Genes that have diverged to encode different but related proteins, such as tyrosinase and TRP-1 (paralogous genes) also often show different but close intron locations (for example, 49, 50). Intron insertion must be invoked to explain the evolution of these genes; otherwise in some cases the ancestral exons would be of only a few base-pairs in length. What explanation can be offered for the finding, as in this study, that non-homologous introns do nevertheless tend to occur in similar regions of homologous genes? In general it is also found that protein encoding exons have a fairly narrow size distribution (51). We suggest that selection for splicable transcripts may in part be the cause. If the RNA splicing machinery requires particular secondary structure folding, in addition to the consensus splice sites small exons might be disfavoured and lost in evolution. The present-day eukaryotic gene structures are the outcome of large-scale random intron insertion followed by selection.

#### Promoter sequences of TRP-1 and tyrosinase

Mouse and human tyrosinase genes have both CCAAT and TATA boxes (25, 26, 29). The former sequence is a core element capable of binding a number of transcription factors. The TATA

box is normally found, as in these two genes, within 25–40 bp 5' of transcriptional initiation and binds TFIID which directs accurate initiation. Most tissue-specific genes contain these *cis*-acting elements. However, although the murine TRP-1 gene promoter does have the sequence motif TATA, it is more than 180 bp upstream of transcriptional initiation and cannot be functioning as a TFIID binding site. The heterogeneous 5' ends of the transcript bear this out. Nor is there is a CCAAT element upstream of TRP-1, but on the opposite strand there is the core of the NF-1 binding site (GCCAAT seen as ATTGGC at bases 142–148 in Figure 2). This 'inverted' CCAAT box has a 7/12 match to the core of the Y-box element (CTGATTGGCCAA) found upstream of both some ubiquitous and some tissue specific genes (52, 53). There are no binding sites for transcription factors Sp1, CREB, AP-1, AP-2, Oct-1 or 2, C/EBP COUP/TF or SRF.

The tyrosinase and TRP-1 genes not only show common ancestry, they are also similarly regulated; both are melanocyte-specific transcripts (7, 20, 24). The lack of obvious similarity between the promoter regions is thus puzzling. If the 5'-untranslated exon has been acquired by TRP-1 during its evolution some of the regulatory sequences may now be located in the first intron; which therefore might have enhancer function. Searching of the intron sequence does not identify any characterised transcription factor binding sites. Furthermore, as the 370–470bp 5' of the gene alone are sufficient to direct specific transcription in the cultured cell assay, specificity must reside in this region.

As little as 270bp of 5' flanking DNA is sufficient to direct tyrosinase expression specifically in cultured cells and in transgenic animals (54). Comparison of the short regions 5' of mouse tyrosinase and TRP-1 reveals only a few sequence motifs present in both. The largest of these is an 11bp sequence, with some palindromic character, AGTCATGTGCT, boxed in Figure 2. This is the only motif which is also found in the human tyrosinase gene. The sequence is located between 40 and 140bp 5' of the start of transcription of all three genes. Occurrence of such a sequence by chance is very unlikely, and we suggest that it has been conserved as a functional element. We propose that it is a factor-binding element responsible, wholly or partly for the melanocyte-specific transcription of the tyrosinase and TRP-1.

#### ACKNOWLEDGEMENTS

We would like to thank Gunther Schutz and Sigi Ruppert for communication of results prior to publication. The work is supported by the Medical Research Council and the Lister Institute for Preventive Medicine. IJJ is a Lister Fellow. The sequence reported in this paper is deposited in the EMBL database, accession number X59513.

#### REFERENCES

- Hearing, V.J. and Jimenez, M. (1989) *Pigment Cell Res.* **2**, 75–85.
- Korner, A.M. and Pawelek, J.M. (1982) *Science*, **217**, 1163–1165.
- Korner, A.M. and Pawelek, J.M. (1980) *J. Invest. Dermatol.*, **75**, 192–195.
- Murray, M., Pawelek, J.M. and Lamoreux, M.L. (1983) *Develop. Biol.*, **100**, 120–126.
- Barber, J.I., Townsend, D., Olds, D.P. and King, R.A. (1984) *J. Invest. Dermatol.*, **83**, 145–149.
- Kwon, B.S., Haq, A.K., Pomerantz, S.H. and Halaban, R. (1987) *Proc. Nat. Acad. Sci. U.S.A.*, **84**, 7473–7477.
- Muller, G., Ruppert, S., Schmid, E. and Schutz, G. (1988) *EMBO J.*, **7**, 2723–2730.

8. Yamamoto, H., Takeuchi, S., Kudo, T., Makino, K., Nakata, A., Shinoda, T. and Takeuchi, T. (1987) *Jpn. J. Genet.*, **62**, 271–274.
9. Kwon, B.S., Wakulchik, M., Haq, A.K., Halaban, R. and Kestler, D. (1988) *Biochem. Biophys. Res. Comm.*, **153**, 1301–1309.
10. Kwon, B.S., Haq, A.K., Wakulchik, M., Kestler, D., Barton, D.E., Franke, U., Lamoreux, M.L., Whitney, J.B. and Halaban, R. (1989) *J. Invest. Dermatol.*, **93**, 589–594.
11. Tomita, Y., Takeda, A., Okinaga, S., Tagami, H. and Shibahara, S. (1989) *Biochem. Biophys. Res. Comm.*, **164**, 990–996.
12. Giebel, L.B., Strunk, K.M., King, R.A., Hanifin, J.M. and Spritz, R.A. (1990) *Proc. Nat. Acad. Sci. U.S.A.*, **87**, 3255–3258.
13. Spritz, R.A., Strunk, K.M., Giebel, L.B. and King, R.A. (1990) *New England J. Med.*, **322**, 1724–1728.
14. Kikuchi, H., Hara, S., Ishiguro, S., Tamai, M. and Watanabe, M. (1990) *Hum. Genet.*, **85**, 123–124.
15. Takeda, A., Tomita, Y., Matsunaga, J., Tagami, H. and Shibahara, S. (1990) *J. Biol. Chem.*, **265**, 17792–17797.
16. Shibahara, S., Okinaga, S., Tomita, Y., Takeda, A., Yamamoto, H., Sato, M. and Takeuchi, T. (1990) *Eur. J. Biochem.*, **189**, 455–461.
17. Jackson, I.J. and Bennett, D.C. (1990) *Proc. Nat. Acad. Sci. U.S.A.*, **87**, 7010–7014.
18. Kwon, B.S., Halaban, R. and Chintamaneni, C. (1989) *Biochem. Biophys. Res. Comm.*, **161**, 252–260.
19. Yokoyama, T., Silversides, D.W., Waymire, K.G., Kwon, B.S., Takeuchi, T. and Overbeek, P.A. (1990) *Nucleic Acid. Res.*, **18**, 7293–7298.
20. Shibahara, S., Tomita, Y., Sakakura, T., Nager, C., Chandhuri, B. and Muller, R. (1986) *Nucleic Acid. Res.*, **14**, 2413–2427.
21. Jackson, I.J. (1988) *Proc. Nat. Acad. Sci. U.S.A.*, **85**, 4392–4396.
22. Zbarsky, E., Favor, J. and Jackson, I.J. (1990) *Genetics*, **126**, 443–449.
23. Rinchik, E.M., Bangham, J.W., Hunsicker, P.R., Cacheiro, N.L.A., Kwon, B.S., Jackson, I.J. and Russell, L.B. (1990) *Proc. Nat. Acad. Sci. U.S.A.*, **87**, 1416–1420.
24. Ruppert, S., Muller, G., Kwon, B. and Schutz, G. (1988) *EMBO J.*, **7**, 2715–2722.
25. Yamamoto, H., Takeuchi, S., Kudo, T., Sato, C. and Takeuchi, T. (1989) *Jpn. J. Genet.*, **64**, 121–135.
26. Kikuchi, H., Miura, H., Yamamoto, H., Takeuchi, T., Dei, T. and Watanabe, M. (1989) *Biochim. Biophys. Acta*, **1009**, 283–286.
27. Tanaka, S., Yamamoto, H., Takeuchi, S. and Takeuchi, T. (1990) *Development*, **108**, 223–227.
28. Beermann, F., Ruppert, S., Hummler, E., Bosch, F.X., Muller, G., Ruther, U. and Schutz, G. (1990) *EMBO J.*, **9**, 2819–2826.
29. Giebel, L.B., Strunk, K.M. and Spritz, R.A. (1991) *Genomics*, **9**, 435–445.
30. Church, G.M. and Gilbert, W. (1984) *Proc. Nat. Acad. Sci. U.S.A.*, **81**, 1991–1995.
31. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
32. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science*, **239**, 487–491.
33. Winship, P.R. (1989) *Nucleic Acid. Res.*, **17**, 1266.
34. Luckow, B. and Schutz, G. (1987) *Nucleic Acid. Res.*, **15**, 5490.
35. Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning; a Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
36. Mount, S.M. (1982) *Nucleic Acid. Res.*, **10**, 459–472.
37. Shapiro, M.B. and Senepathy, P. (1986) *Nucleic Acid. Res.*, **15**, 7155–7174.
38. Senepathy, P., Shapiro, M.B. and Harris, N.L. (1990) *Methods Enzymol.*, **183**, 252–278.
39. Aebi, M., Hornig, H. and Weissmann, C. (1987) *Cell*, **50**, 237–246.
40. Huber, M., Hinterman, G. and Lerch, K. (1985) *Biochemistry*, **24**, 6038–6044.
41. Lerch, K., Longoni, C. and Jordi, E. (1982) *J. Biol. Chem.*, **257**, 6408–6413.
42. Cavalier-Smith, T. (1991) *Trends Genet.*, **7**, 145–148.
43. Rogers, J.H. (1985) *Nature*, **315**, 458–459.
44. Rogers, J.H. (1986) *Trends Genet.*, **2**, 223.
45. Rogers, J.H. (1989) *Trends Genet.*, **5**, 213–216.
46. Rogers, J.H. (1990) *FEBS Lett.*, **268**, 339–343.
47. Dibb, N.J. and Newman, A.J. (1989) *EMBO J.*, **8**, 2015–2021.
48. Chen, C.-J., Clark, D., Ueda, K., Pastan, I., Gottesman, M.M. and Roninson, I.B. (1990) *J. Biol. Chem.*, **265**, 506–514.
49. Ali, S. and Clark, A.J. (1988) *J. Mol. Biol.*, **199**, 415–426.
50. Kirchgessner, T.G., Chuat, J.-C., Heinzmann, C., Etienne, J., Guilhot, S., Svenson, K., Ameis, D., Pilon, C., d'Auriol, L., Andalibi, A., Schotz, M.C., Galibert, F. and Lusi, A.J. (1989) *Proc. Nat. Acad. Sci. U.S.A.*, **86**, 9647–9651.
51. Traut, T.W. (1988) *Proc. Nat. Acad. Sci. U.S.A.*, **85**, 2944–2948.
52. Dider, D.K., Schiffenbauer, J., Woulfe, S.L., Zacheis, M. and Schwartz, B.D. (1988) *Proc. Nat. Acad. Sci. U.S.A.*, **85**, 7322–7326.
53. Tafuri, S.R. and Wolffe, A.P. (1990) *Proc. Nat. Acad. Sci. U.S.A.*, **87**, 9028–9032.
54. Kluppel, M., Beerman, F., Ruppert, S., Schnid, E., Hummler, E. and Schutz, G. (1991) *Proc. Nat. Acad. Sci. U.S.A.*, **88**, 3777–3781.
55. Jackson, I.J. (1991) *Nucleic Acid. Res.*, **19**, 3795–3798.